

# SCIENTIFIC REPORTS

OPEN

## Metagenomics of pigmented and cholesterol gallstones: the putative role of bacteria

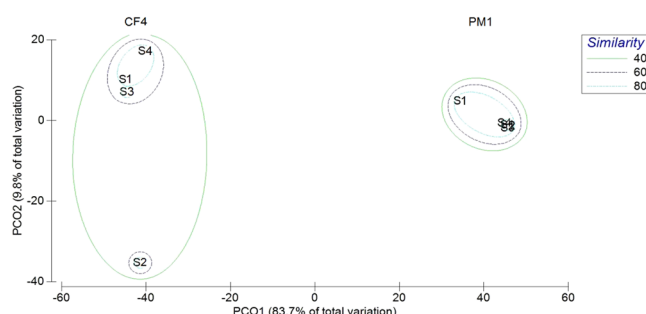
S. H. Kose<sup>1</sup>, K. Grice<sup>1</sup>, W. D. Orsi<sup>2</sup>, M. Ballal<sup>3,4</sup> & M. J. L. Coolen<sup>1</sup>

There is growing evidence for bacteria playing a role in the pathogenesis and formation of pigmented gallstones from humans. These studies mainly involved cultivation of gallstone-associated bacteria and 16S rRNA profiling, providing an indirect link between processes involved in gallstone formation by the bacteria *in-situ*. Here, we provide functional metagenomic evidence of a range of genes involved in bile stress response, biofilm formation, and anaerobic energy metabolism by Gram-negative *Klebsiella* in pigmented gallstones from a 76-year-old male patient. *Klebsiella* was also present in one cholesterol-type stone in a 30-year-old female patient who had additional cholesterol gallstones characterised by Gram-positive bacteria. Pigmented stones further revealed a predominance of genes involved in carbohydrate metabolism, whilst cholesterol stones indicated a profile dominated by protein metabolism possibly reflecting known chemical differences between Gram-negative and Gram-positive biofilm matrices. Archaeal genes were not detected. Complementary carbon and hydrogen isotopic analyses of cholesterol within the patients' stones revealed homogeneity, suggesting a common diet or cholesterol biosynthesis pathway that has little influence on microbial composition. This pilot study provides a framework to study microbial processes that play a potential role in gallstone formation across markedly different types of stones and patient backgrounds.

The focus on bacteria and its role in gallstone pathogenesis began most notably in 1966 by Maki<sup>1</sup>, and furthered by Stewart *et al.*<sup>2,3</sup>. In those studies, bacteria were suggested to play a causal role in the pathogenesis of pigmented and the pigmented portion of mixed stones only, as cholesterol stones rarely exhibited bacterial signatures<sup>2</sup>. A conclusive definition of gallstone types and their bacterial compositions are yet to be determined due to the complex nature of gallstones. However, researchers do generally group the stone types into cholesterol (predominantly composed of cholesterol, that may have pigmented centres), pigmented (predominantly composed of the bile pigment bilirubin, derived from the break down of aged red blood cells by the liver giving its brown colour), mixed (a compositional mixture of the cholesterol and pigmented) and black (assumed to be pigmented but with a black 'volcanic glass' type appearance). For a detailed discussion on stone types and compositions see Stewart *et al.*<sup>4</sup>. A further study restricted gallstone pathogenesis *via* bacterial action to only brown pigmented stones, ruling out cholesterol, mixed pigmented and black pigmented gallstones altogether<sup>5</sup>. Subsequently, bacterial studies were focused on pigmented stone formation with the dominant theory behind the mechanism of formation for this type of stone through the activity of bacterially produced  $\beta$ -glucuronidase<sup>1</sup>.  $\beta$ -glucuronidase was observed to be the deconjugating factor that led to the precipitation of calcium bilirubinate crystals, with these crystals conjugated by an anionic glycoprotein (i.e. sodium alginate) leading to the agglomeration of the bilirubinate calcium crystals into macroscopic stones<sup>1</sup>. However, Maki's findings related to *Escherichia coli* (a known producer of  $\beta$ -glucuronidase), and did not explore non- $\beta$ -glucuronidase producing bacterial species that are often found in gallstone studies<sup>7</sup>. Biofilm formation as an alternative mechanism for the role of bacteria in pigmented stone formation was initially proposed by Stewart *et al.*<sup>2</sup>. In that culture-based study, the biofilm product glycocalyx (an anionic glycoprotein), was suggested to be the central agglomerating factor, with  $\beta$ -glucuronidase having a comparatively minor role. *Klebsiella*, *Enterococcus*, *Enterobacter*, *E. coli* and *Psuedomonas aeruginosa*, were reported to be the most prevalent cultured bacteria across 61 stones (predominantly pigmented and mixed stones), with

<sup>1</sup>School of Molecular and Life Sciences, W.A Organic and Isotopic Geochemistry Centre, Curtin University, Perth, WA, 6000, Australia. <sup>2</sup>Department of Earth and Environmental Science, GeoBio Centre LMU, Ludwig-Maximilians-Universität München, 80333, Munich, Germany. <sup>3</sup>Fiona Stanley Hospital, 11 Robin Warren Dr, Murdoch, 6150, WA, Australia. <sup>4</sup>St John of God Murdoch Hospital, Barry Marshall Parade, Murdoch, 6150, WA, Australia. Correspondence and requests for materials should be addressed to S.H.K. (email: [sureyya.kose@curtin.edu.au](mailto:sureyya.kose@curtin.edu.au))

SampleID	Age	Sex	BMI	Antibiotics	Type	Size	No.	Site
PM1	76	M	24	Nil	Pigmented	Small 0.5–1 cm	4	FSH
CF4	30	F	40.5	Nil	Cholesterol	Small 0.5–1 cm	4	SJOG

**Table 1.** Patient Information.**Figure 1.** Principle coordinate analysis (PCoA) ordination of Bray-Curtis similarity (square root transformed) between total bacterial genera in the four gallstones of patients PM1 and CF4. Shown is the first two principle coordinate axes, which combined explain 93.5% of the variation between the patients. Coloured ellipses signify the percentage of similarity between the patients' native stones.

*P. aeruginosa* the greatest biofilm producer of the group after *Citrobacter freundii*<sup>3</sup>. A more recent study involving mice as a host identified *Salmonella*, via scanning electron microscopy (SEM) and culture analysis, as a primary producer of biofilms on the surface of cholesterol gallstones<sup>6</sup>. However, only up to 1% of bacteria in complex environmental samples can generally be brought into culture and provides a biased view into the relative abundance of the total species present. A recent cultivation-independent investigation of the microbial composition of 27 cholesterol gallstones using high-throughput 16S rRNA profiling identified a predominance of *Enterobacteriaceae* and *Ruminococcaceae* and to a lesser extent, *Bacteroidales*, *Lactococcus*, *Enterococcus* and *Clostridiales* within the analysed gallstones<sup>7</sup>. However, 16S rRNA profiling does not provide information on the role the identified bacteria play in the formation of gallstones and what mechanisms they possess to survive in the human gallbladder.

In this study we analysed functional metagenomes<sup>8</sup> to investigate the diversity and metabolic potential of microbial communities in pigmented vs. cholesterol stones, and whether they possess genes involved in the formation of biofilms or other processes including bile resistance that could lead to gallstone formation. The sequencing of functional metagenomes was furthermore used to provide parallel information on the functional and/or taxonomic diversity of all domains of life not limited to bacteria (e.g. fungi, archaea, and viruses) that may be present and underexplored in gallstones<sup>9,10</sup>. To the best of our knowledge, the only other related shotgun metagenomics study was performed on human bile samples<sup>11</sup>. However, it has been shown that bacterial biofilms may persist on the surface of gallstones, even when the patient's bile is culture negative<sup>12</sup>.

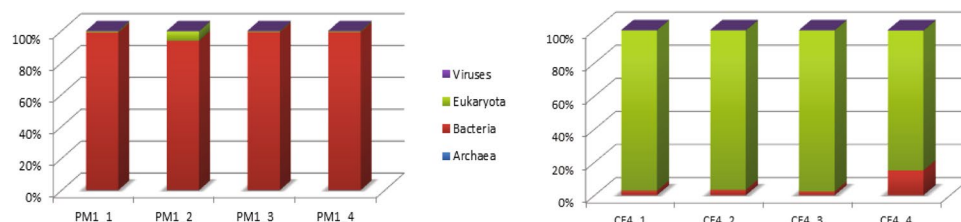
We complement the shotgun metagenomics analysis with compound specific isotopic analysis (CSIA) of cholesterol in the gallstones - namely carbon ( $\delta^{13}\text{C}$ ) and hydrogen ( $\delta\text{D}$ ) isotopes to ascertain possible dietary or exogenous environmental factors that may be associated with or divergent from the bacteria identified in this study.

## Results and Discussion

Gallstones were collected from the gallbladders of a 76-year old male patient (PM1) with pigmented type gallstones ( $n = 4$ ) and one 30-year old female patient (CF4) with cholesterol type gallstones ( $n = 4$ ) (Table 1). Both patients were diagnosed with gallstone disease (cholelithiasis) by ultrasound imaging and computed tomography scanning at the Fiona Stanley Hospital and St John of God Hospital, Murdoch, WA, Australia. Neither patient had been given a course of antibiotics for treatment. Pigmented gallstones contained  $57.2 \pm 36.6$  ng per stone genomic DNA while the DNA content in the cholesterol gallstones was  $16.5 \pm 10.9$  ng per stone. This DNA served as template for the construction of the metagenomics libraries discussed below.

**Metagenomics.** Metagenomic analysis revealed distinct differences between the microbial diversity of the two patients. The following section focuses on the taxonomic diversity and gene functions potentially associated with gallstone pathogenesis and formation.

Principle component analysis (PCoA; Fig. 1) on open reading frames (ORFs) revealed that the first component explains 83.7% of the variance in microbial diversity in the analysed gallstones. The second component explains 9.8% of the variation, with both components combined accounting for 93.5% of the variation between the two patients. Only 5.5% of the variation remains unexplained. The microbial composition between the four stones found in patient PM1 showed a high level of similarity at 80%. A 60% similarity in microbial diversity was observed between stones 1, 3 and 4 of patient CF4. Stone 2, however, only shared 40% similarity to the group (Fig. 1).



**Figure 2.** Relative abundance of bacteria, archaea, eukaryotes and viruses recovered from metagenomes in patients PM1 (left) and CF4 (right).

**Taxonomic affiliation.** Taxonomic analysis of the ORFs of 4 replicate pigmented gallstones of patient PM1 revealed a predominance of bacteria (97.65%) while no archaeal ORFs were recovered (Fig. 2). Eukaryotes comprised 1.98% of the ORFs with the majority (80.49%) of human origin. A small proportion of the ORFs were of viral origin (0.3%) and misidentified reads (0.07%).

Taxonomic analysis of bacterial ORFs comprising >1% of the total community revealed that Gram-negative *Klebsiella* was the most abundant genus in PM1 (69–79%), followed by *Enterococcus* (3.7–13.6%), *Escherichia* (2.9–8.0%), *Salmonella* (1.7–2.2%), and *Enterobacter* (1.1–1.3%). *Meiothermus* was abundant in stone 1 (17.4%), but was not detected in the other stones of PM1 (Fig. 3).

Metagenomic profiling of 4 replicate cholesterol gallstones of patient CF4 revealed a predominance of eukaryotes (96.05%) with the majority (98.5%) stemming from the patient or comprising human cells, with others (1.46%). Bacteria comprised only 3.6% of the ORFs. In agreement with the metagenomes of PM1 stones, a small portion of the ORFs were of viral origin (0.3%) and misidentified reads (0.05%), while no archaeal ORFs were recovered (Fig. 2).

Taxonomic analysis of bacterial ORFs comprising >1% of the total bacterial community revealed that Gram-positive *Streptococcus* was the most abundant genus within all 4 stones in CF4 (38.4–54.7%) (Fig. 3), followed by *Alconivorax* (6.4–27.8%), *Mollicutes* (6.3–15.1%), *Nocardia* (7.0–13.0%), *Methanoculleus* (5.5–8.1%), *Rickettsia* (4.9–6.9%), and *Cystobacter* (2.7–4.1%). Stone 2 only showed 40% similarity with the other three stones as it did not contain *Nocardia*, *Methanoculleus*, *Rickettsia*, and *Cystobacter*. It was, however, the only stone to exhibit *Klebsiella* (6.8%) and other *Enterobacteriaceae* (4.5%). *Geobacillus* was found only in stones 3 and 4 (1.1–1.2%).

As mentioned earlier, a consensus is yet to be drawn on what type of bacterial communities are common to pigmented and cholesterol gallstones<sup>7</sup>. This is in part due to the complex and unique interplay between a patient's health history, microbiome, environment, and a predominance of culture studies in which total community resolution is limited<sup>7</sup>. Nevertheless, in our study, patient PM1's pigmented stone bacterial community do approximately align with a previous culture analysis in which *Klebsiella*, *Enterococcus*, *Enterobacter*, *E. coli* and *Pseudomonas aeruginosa* were found to be the most prevalent genera across 61 pigmented and mixed gallstones<sup>3</sup>.

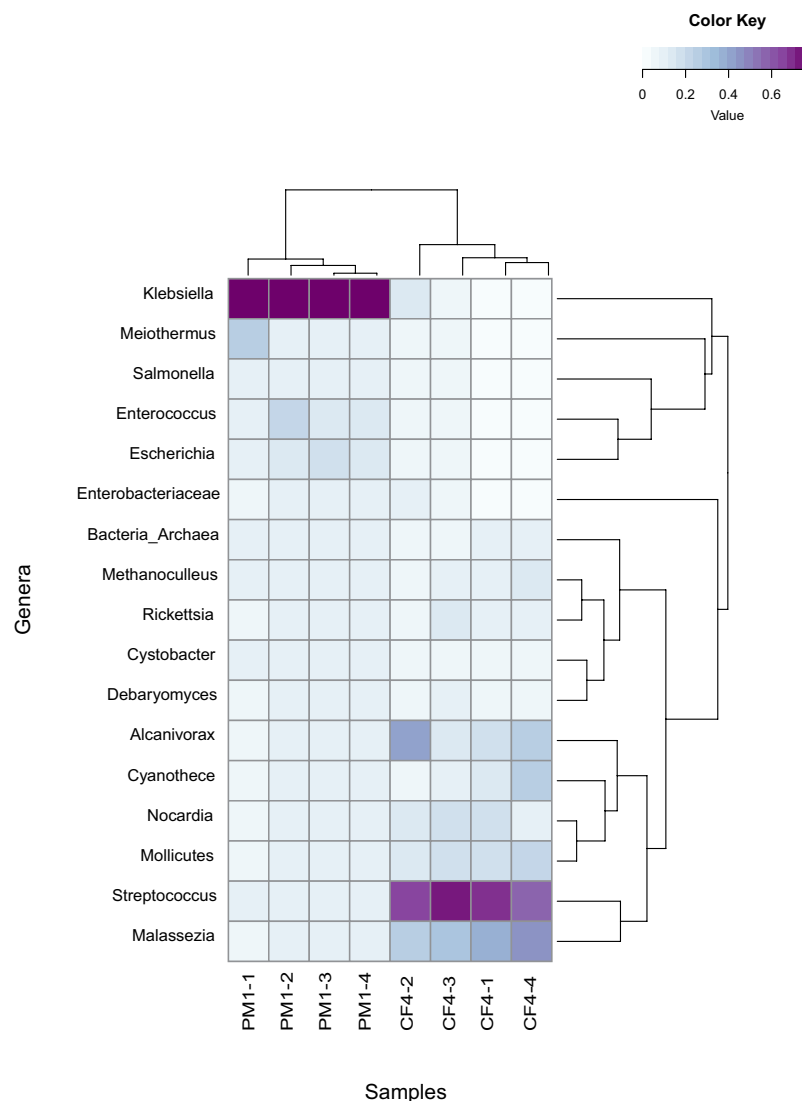
Less is known about microbial compositions in cholesterol gallstones, due to their propensity to have low amounts of bacterial DNA. However, this assertion may be due to the limitation of culture studies, as a recent investigation of the microbial composition of 27 cholesterol gallstones using high-throughput 16S rRNA profiling identified *Enterobacteriaceae* to be abundant<sup>7</sup>, whereas the less abundant genera differed from those present in patient CF4 in our study.

Quantitative PCR revealed a two orders of magnitude higher bacterial 16S rRNA copy number ( $\sim 9.9 \times 10^6 \pm 4.6 \times 10^6$ ) in pigmented gallstones of PM1 compared to the cholesterol-type gallstones of CF4 ( $3.1 \times 10^4 \pm 2.0 \times 10^4$  copies per stone). The results of this cultivation-independent molecular quantification approach are consistent with previous culture studies that report a greater microbial biomass in pigmented vs. cholesterol stones<sup>1–3,13</sup>. We provide a conservative estimate of bacterial cell numbers ( $1.2 \times 10^6 \pm 5.8 \times 10^5$ ) bacterial cells per stone in PM1 and  $7.7 \times 10^3 \pm 5.0 \times 10^3$  cells per stone in CF4) based on known number of ribosomal operons in the most abundant genera found in PM1 (*Klebsiella* with 8 rRNA operons<sup>14</sup>) and CF4 (*Streptococcus* with 4 rRNA operons<sup>15</sup>).

**Functional metagenomic profiling.** *Genes involved in Bile Stress response.* Microorganisms have been known to survive and thrive in a range of extreme environments, including the human body where variations in pH, nutrient limitations, low oxygen levels, an established diverse bacterial microbiome, and host immunity responses provide longterm obstacles for survival<sup>16–18</sup>.

The human gallbladder environment poses its own unique challenges. The liver secretes approximately 800–1000 mL of bile into the gallbladder per day<sup>19</sup>. Bile acts as a detergent or emulsification agent for the digestion and absorption of lipids and contains sodium and potassium salts, bile acids (namely chenodeoxycholic acid, cholic acid), cholesterol, phospholipids, and bile pigments such as bilirubin<sup>19</sup>. This environment is made further hostile to bacteria as bile is concentrated 5–10 fold in the gallbladder, with commonly used antibiotic drugs, and heavy metals being secreted into bile as per the liver's detoxification and enterohepatic cycling processes<sup>16,19</sup>.

The strategies bacteria employ to resist toxic agents such as bile and antibiotics are often complex and involve a variety of methods that include and are not limited to efflux pumps (which pump the toxic agent out of the cell), reduction of cell permeability, enzymatic modification or destruction of toxic agents either within or without the cell wall barrier, and the modification of the toxic agent's target either *via* genetic mutation, enzymatically, or by presenting an alternate target<sup>20,21</sup>. Bile, as a detergent, and consisting of a variety of toxic agents has been shown



**Figure 3.** Heatmap with the major bacterial genera identified in the gallstones ( $n = 4$ ) of patients PM1 (pigmented) and CF4 (cholesterol). The color key shows the relative abundance of the genera in the gallstones. The dendrograms illustrate the relationship between samples showing that the distribution of genera is relatively similar between replicate stones, but greatly differ between the two patients.

to cause membrane perturbations, DNA damage and oxidative stress in bacteria<sup>21</sup>. This is shown to be consistent with our current study where multidrug export efflux pumps, DNA and cell wall repair proteins were identified as important in bile resistance. Of the seventeen taxa in both patients combined that comprised at least 1% of the total sequenced gene pool, we identified six genera that harbour genes involved in bile stress survival, which are linked to the production of biofilms that may be associated with pigmented as well as cholesterol-type gallstone formation (Tables 2, 3 and 4) as discussed in detail below.

**Gram-negative Bacteria.** *Klebsiella* (bile-sensitive genes). The *marABC* and *marR* operons (for all genes identified see Table 2) are regulatory genes that control multiple antibiotic drug resistance<sup>22</sup> and have been shown to be activated in the presence of the bile salt deoxycholate, with the level of gene expression dependent on the salts' concentration<sup>16,23</sup>. The *Tol* protein and derivatives are important in many Gram-negative bacterial transport systems and act as an outer membrane pore function or efflux pump<sup>24</sup>. Mutations in *Tol* genes destabilise the membrane allowing for greater bile salt entry, thereby affecting bile resistance<sup>24</sup>. The *TolC* efflux pump (Table 2), in particular, has been shown to be upregulated in biofilms and is associated with the removal of toxic compounds and antibiotics<sup>21</sup>. The *emrEB*, *mdtABCD*, *cmeAB* genes similarly correspond to efflux pump action and are essential for bile resistance<sup>24</sup>. The *emrEB* multidrug efflux pump systems have been shown to actively efflux the bile salt chenodeoxycholic acid<sup>25</sup>. Furthermore, over-expression of *mdtABCD*, a multidrug resistance efflux pump cluster, leads to increased deoxycholate resistance<sup>26</sup>. The *cmeAB* has been shown to function as a multidrug efflux pump in *C. jejuni* by effectively mediating resistance to bile salts<sup>27</sup>. The *Dam* (DNA adenine methylase) enzyme has been associated with repairing damage to *Salmonella* DNA after bile acid exposure<sup>28</sup>. Similarly, the *sbcC* and *MutS*

Genes disrupted in bile-sensitive mutants	Function of gene products/putative function	Reference(s)
<b>Gram-negative bacteria</b>		
<i>Klebsiella</i>		
<i>marA</i>	Multiple antibiotic resistance protein – Regulatory genes	11,19,20
<i>marB</i>	Multiple antibiotic resistance protein – Regulatory genes	11,19,20
<i>marC</i>	Multiple antibiotic resistance protein – Regulatory genes	11,19,20
<i>marR</i>	Regulatory genes	11,19,20
<i>TolC precursor</i>	Efflux pump function	17,21
<i>dam</i>	DNA adenine methylase	24
<i>emrE</i>	Efflux pump – drug resistance	31
<i>emrB</i>	Efflux pump – drug resistance	31
<i>mdtABCD</i>	Efflux pump – multi drug resistance	22,31
<i>cmeA</i>	Efflux pump	23,31
<i>cmeB</i>	Efflux pump	23,31
<i>sbcC</i>	Exonuclease	25
<i>MutS</i>	DNA mismatch repair	25
<i>nifH</i> homolog	Pyruvate flavodoxin oxidoreductase	25
<i>yvaG</i>	3-oxoacyl-acyl-carrier protein reductase	25
<i>dgt</i> homolog	Deoxyguanosinetriphosphate triphosphohydrolase	25
<i>Escherichia</i>		
<i>yvaG</i>	3-oxoacyl-acyl-carrier protein reductase	25
<i>Shigella</i>		
<i>PhoQ</i>	Regulatory genes	34
<b>Gram-positive bacteria</b>		
<i>Enterococcus</i>		
<i>emrB</i>	Efflux pump – drug resistance	31
<i>sbcC</i> homolog	Exonuclease	25
<i>MutS</i>	DNA mismatch repair	25
<i>nifH</i> homolog	Pyruvate-flavodoxin oxidoreductase	25
<i>yvaG</i>	3-oxoacyl-acyl-carrier protein reductase	25
<i>dgt</i> homolog	Deoxyguanosinetriphosphate triphosphohydrolase	25

**Table 2.** Loci disrupted in bile-sensitive mutants and the functions of gene products.

(DNA repair), *yvaG* (rebuilding the cell membrane after stress), *nifH* (oxidative response) and *dgt* (dGTP hydrolysis) homologs were associated with DNA and cell wall repair in response to bile stress in *Enterococcus faecalis*<sup>29</sup>.

*Klebsiella* (ORFs, promoters, proteins induced by bile). Experiments with bile salt treatments to *Enterococcus faecalis*<sup>16,29</sup> identified an increased production of a number of stress proteins (*Gsp*). Three of these stress proteins were identified in this study as the *DnaK* and *GroEL/GroES* molecular chaperones<sup>29</sup> and the organic hydroperoxide resistance protein *Ohr*<sup>16,30</sup> within *Klebsiella* (Table 3). Further experiments with *Propionibacterium freudenreichii* revealed genes involved in a variety of stress responses (heat, acid, bile salts) termed *GSPs* (General stress proteins), with bile salts in particular associated with oxidative stress responses<sup>31</sup>. A number of these were annotated to *Klebsiella* in this study and include the molecular chaperones *Hsp20* (heat stress), *DnaK*, *GoEL*, *AspA* and *ClpB* that are associated with acid stresses<sup>31</sup>. *SodA*, an oxidative damage remediation gene, was also identified in this study and has been shown to be involved in stress responses within *Lactobacillus lactis* (oxygen stress), *Bacillus subtilis* (heat, salt and ethanol stresses), and *B. cereus* (heat, salt and ethanol stress)<sup>31,32</sup>. Further oxidative damage reduction and remediation proteins identified for *Klebsiella* were *ORF002*, *G6PD*, and the *CysK*, *HemH* homologs. The NADPH dependent aldo or keto-oxidoreductase *ORF002* protein is an important part of the glutathione cellular defense system that is involved in the reduction of oxidative stress caused by reactive oxygen species (ROS) associated with bile<sup>32,33</sup>. The *G6PD* (Glucose-6-phosphate 1-dehydrogenase) protein has been shown to be activated in the presence of ROS, that arise due to stresses such as high levels of salt, and considered vital for cellular redox balance<sup>34,35</sup>. Similarly, the *CysK* (cysteine synthase) and *HemH* (ferrochelatase) homologs have also been shown to be overexpressed when exposed to bile-salt stresses<sup>31,36</sup>. Other acid stress proteins identified were *BCCP* (a biotin containing carboxyl carrier protein) and *ATPG* (ATP synthase gamma chain)<sup>31,37</sup>. The DNA damage repair proteins *MutB* (Methylmalonyl-CoA mutase) and *RecR* (Recombinase) were also identified<sup>16,36</sup>.

Genes encoding the dissimilatory dimethylsulfide reductase A,B and C (*dmsABC*, Table 3) were retrieved in *Klebsiella*, indicating its capacity for anaerobic metabolism and use of dimethylsulfoxide (DMSO) as a terminal electron acceptor. DMSO respiration is energetically favourable under anaerobic conditions in bacteria that contain this metabolic potential<sup>38</sup>. Further, the *dmsABC* operons are controlled by the oxidative regulator *fnr*, important for oxidative stress response and anaerobic metabolism in pathogenic bacteria<sup>39–41</sup>. This ability by *Klebsiella*

ORFs/promoters/proteins induced by bile	Function of gene products/putative function	Reference(s)
<b>Gram-negative bacteria</b>		
<i>Klebsiella</i>		
<i>DnaK</i>	Molecular chaperone	18,25
<i>GroEL/GroES</i>	Heat shock protein- Molecular chaperone	18,25
<i>Ohr</i>	Organic hydroperoxide resistance	11,18,25
<i>Hsp20</i>	Heat shock protein, molecular chaperone	18
<i>ClpB</i>	ATP-binding chain of an ATP-dependent protease	18
<i>RecR homolog</i>	Recombinase (DNA repair)	11,33
<i>sodA</i>	Manganese superoxide dismutase	18,27
<i>BCCP</i>	Biotin-containing carboxyl carrier protein of acetyl-CoA carboxylase	18,57
<i>CysK homolog</i>	Putative cysteine synthase	18,33
<i>ORF002</i>	NADPH dependent aldo or keto-oxidoreductase	28,29
<i>MutB</i>	Putative Methylmalonyl-CoA mutase	11,33
<i>AspA</i>	Putative Aspartate ammonia-lyase	18
<i>G6PD</i>	Glucose-6-phosphate 1-dehydrogenase	30,32
<i>ATPG</i>	ATP synthase gamma chain	18,57
<i>HemH homolog</i>	Putative Ferrochelatase, protoheme ferro-lyase	18,33
<i>dmsABC</i>	Anaerobic dimethyl sulfoxide reductase chain A,B & C	39–41
<i>Escherichia</i>		
<i>ORF002</i>	NADPH dependent aldo or keto-oxidoreductase	30,32
<i>Shigella</i>		
<i>ORF001</i>	NADPH dependent aldo or keto-oxidoreductase	30,32
<i>ORF002</i>	NADPH dependent aldo or keto-oxidoreductase	30,32
<i>Serratia</i>		
<i>ClpB</i>	ATP Binding Chain on ATP-dependent protease	18
<b>Gram-positive bacteria</b>		
<i>Enterococcus</i>		
<i>DnaK</i>	Molecular chaperone	18,25
<i>GroEL/GroES</i>	Heat shock protein- Molecular chaperone	18,25
<i>Ohr</i>	Organic hydroperoxide resistance	11,26
<i>Gsp</i>	General stress protein	11,18,25
<i>clpB</i>	ATP Binding Chain on ATP-dependent protease	18
<i>sodA</i>	Manganese superoxide dismutase	18,27
<i>BCCP</i>	Biotin-containing carboxyl carrier protein of acetyl-CoA carboxylase	18,57
<i>CysK homolog</i>	Putative cysteine synthase	18,33
<i>OppD homolog</i>	Oligopeptide transport ATP-binding protein	18,35
<i>G6PD</i>	Glucose-6-phosphate 1-dehydrogenase	30,32
<i>ATPG</i>	ATP synthase gamma chain	18,57
<i>Bacillus</i>		
<i>OppD homolog</i>	Oligopeptide transport ATP-binding protein	18,35

**Table 3.** Promoters, proteins or open reading frames (ORFs) induced by bile and their functions.

may explain its successful survival and growth in the anoxic conditions present in the human gallbladder and its dominance in the present study. Other gram-negative and gram-positive bacteria identified in this study may be out competed by *Klebsiella* or utilise fermentation for energy metabolism instead, a less efficient form of energy conservation than DMSO<sup>42</sup>.

*Escherichia (bile-sensitive genes).* Similar to *Klebsiella*, *Escherichia* exhibited the oxido reductase gene *yvaG* or membrane composition and repair protein in our study (Table 2).

*Escherichia (ORF's, promoters, proteins induced by bile).* Like *Klebsiella*, *Escherichia* exhibited the NADPH dependent aldo or keto-oxidoreductase *ORF002* protein involved in the reduction of oxidative stress caused by ROS associated with bile (Table 3).

*Shigella (bile-sensitive genes).* *Shigella* exhibited the *PhoQ* regulatory protein (Table 2). *PhoQ* is closely associated with the *PhoP* regulon. The combined *PhoP-PhoQ* proteins have been associated with various bacteria and their



EPS related genes	Function of gene products/putative function	Reference(s)
<b>Gram-negative bacteria</b>		
<i>Klebsiella</i>		
CsgD	Transcriptional regulator	39
Gsp	General stress protein 18	38
fim Type I	Type 1 fimbriae fimA,B,D,E,L,F,G	12,39
fim Type IV	Type IV fimbrial assembly, ATPase PilB	12,39
wza	Polysaccharide export lipoprotein	12,39
wzc	Tyrosine-protein kinase	12,39
Ribose ABC Transport System	Ribose ABC transport system, ATP-binding protein RbsA	39
Ribose ABC Transport System	Ribose ABC transport system, permease protein RbsC	39
Autoinducer 2(AI-2)	ABC transport system, fused AI2 transporter subunits and ATP-binding component	39
CP4-57- integrase	putative CP4-57-type integrase	39
Polyphosphate kinase	Polyphosphate kinase- Biofilm development	48
sugE	Quaternary ammonium compound-resistance protein	12
ClpX	ATP-dependent Clp protease ATP-binding subunit	12
RapA	RNA polymerase associated protein reg. yhcQ, YeeZ	17
LuxR	Transcriptional regulator	12
CspD	Cold shock protein CspD	12
<i>Escherichia</i>		
wzb	Low molecular weight protein-tyrosine-phosphatase	39
wzc	Tyrosine-protein kinase	39
<b>Gram-positive bacteria</b>		
<i>Enterococcus</i>		
RbsA	Ribose ABC transport system, ATP-binding protein	39
galE	UDP-glucose 4-epimerase	41
sugE	Quaternary ammonium compound-resistance protein SugE	12
ClpX	ATP-dependent Clp protease ATP-binding subunit	12
LuxR	Transcriptional regulator	12
CspD	Cold shock protein CspD	12

**Table 4.** Genes associated with biofilm production and their functions/putative functions.

ability to sense and resist bile stress<sup>43</sup>. Bacterial mutants missing *PhoP-PhoQ* were killed at significantly lower concentrations of bile than those with these proteins, and those with *PhoP* alone surviving a >60% concentration of bile in lab conditions<sup>43</sup>.

*Shigella* (ORF's, promoters, proteins induced by bile). Similar to *Klebsiella* and *Escherichia*, *Shigella* exhibited the NADPH dependent aldo or keto-oxidoreductase *ORF001* and *ORF002* proteins involved in bile oxidative stress reduction (Table 3).

*Serratia* (ORF's, promoters, proteins induced by bile). Bile sensitive genes for *Serratia* were not identified. However, *Serratia* exhibited the *ClpB* molecular chaperone involved in acid stress responses (see *Klebsiella* above).

**Gram-positive bacteria.** *Enterococcus* (bile sensitive genes). Similarly to *Klebsiella*, the important proteins for bile resistance in *Enterococcus* were the multidrug export efflux pump system *emrB*, and those associated with DNA and cell wall repair; *sbsC* homolog, *MutS*, *yvaG*, *nifj* and *dgt* homologs.

*Enterococcus* (ORF's, promoters, proteins induced by bile). As per *Klebsiella*, we see the following same bile stress, oxidative stress and DNA repair genes associated with *Enterococcus*: *Dnak*, *GroEL/GroES*, *Ohr*, *Gsp*, *clpB*, *sodA*, *BCCP*, *G6PD*, *ATPG* and the *CysK* homolog. The *OppD* homolog (ABC transporter – ATP binding protein), involved in the efflux of bile and antibiotic resistance during biofilm formation, was also present<sup>31,44</sup>.

*Bacillus* (ORF's, promoters, proteins induced by bile). Bile sensitive genes for *Bacillus* were not identified in the metagenomes. However, *Bacillus* exhibited the *OppD* homolog as described above for *Enterococcus*.

**Genes involved in biofilm production.** In addition to mediating toxic substances via the resistance strategies described above, microorganisms can group together, attach to either living or non-living surfaces, and form what is called a biofilm<sup>45</sup>. Biofilms comprise a variety of microorganisms enclosed in an extracellular polymeric substance (EPS) or matrix made up of mostly polysaccharides and other environmental specific materials<sup>45</sup>. Biofilm formation has considerable advantages and has shown to protect bacterial communities from UV light, heavy metals, acidity, hydration or salinity changes and host immune responses, including large doses of

antimicrobial drugs that would be lethal to the same community if in a planktonic state<sup>21,46</sup>. There are three main stages involved in biofilm formation: initial adherence to a surface, development of a community structure and ecosystem, and eventual detachment<sup>45</sup>. Each stage is a complex process regulated by a variety of genes that are often environmental, bacterial strain or stressor specific<sup>21</sup>. Common mechanisms include the development of curli fimbriae (adherence or attachment mechanisms), and quorum-sensing or cross community communication to coordinate biofilm attachment, development, detachment, and resistance<sup>13,46</sup>. The main resistance mechanisms afforded by the EPS matrix include drug indifference, in which the EPS works as a barrier between the drug and the targeted microbial cell membrane, the allowance of antibiotics to slowly diffused through the EPS, so that time is given for resistance mutations to develop, efflux pumps, and the secretion of periplasmic glutans that keep toxic substances away from intracellular targets<sup>21</sup>. Certain organisms may provide the base biofilm, whilst others live either competitively or symbiotically within it, with environmental and community composition in a state of constant change<sup>46</sup>. As previously reported, biofilms with *Salmonella* were identified through SEM and culture analysis, on the surface of cholesterol gallstones<sup>6</sup>. Our study also suggested the presence of a biofilm on the surface of both pigmented and cholesterol type gallstones since stringent UV sterilisation initially resulted in a reduction of the yield of extracted genomic DNA by over 90%. Indeed the presence of a microbial biofilm on the surface of the patients' gallstones was confirmed by taxonomic and functional metagenomics analysis. The following genera were identified in the gallstone metagenomes that harbor genes which are putatively associated to biofilm production.

***Klebsiella*.** The *CsgD* gene is a transcriptional activator involved in the regulation of curli fimbriae biosynthesis<sup>13</sup> (Table 4). Curli fimbriae have been identified to be significant EPS components within *Enterobacteriaceae* and involved in bacterial adherence to abiotic substances and cell adhesion during symbiotic or infectious processes<sup>13</sup>. *Klebsiella*, an enterobacterium, has exhibited the *CsgD* gene, alongside the Type 1 and Type IV fimbriae in this study. The Type 1 and Type IV fimbriae are part of the gene cluster *fim*, containing all the components required for fimbrial assembly, and associated with capsule and pilin processes - significant factors in colonisation and biofilm production<sup>17</sup>. Type 1 and Type IV fimbriae have also been shown to facilitate biofilm assembly on both abiotic and host-derived extracellular matrix protein surfaces<sup>13</sup>. The *wza* and *wzc* genes encode for surface molecules involved in capsule assembly and are considered to be important in the early stages of biofilm formation by *Klebsiella pneumoniae*<sup>13,17</sup>. A study isolating the genes involved in biofilm formation of the *K. pneumoniae* strain causing Pyogenic Liver Abscess found *SugE* an important gene that affects biofilm production by modulating capsular polysaccharide production and biofilm mucoviscosity<sup>17</sup>. The *ClpX* and *LuxR* regulatory genes, part of the sugar-specific phosphotransferase systems, and the cold shock protein *cspD* were also implicated in biofilm production<sup>17</sup>.

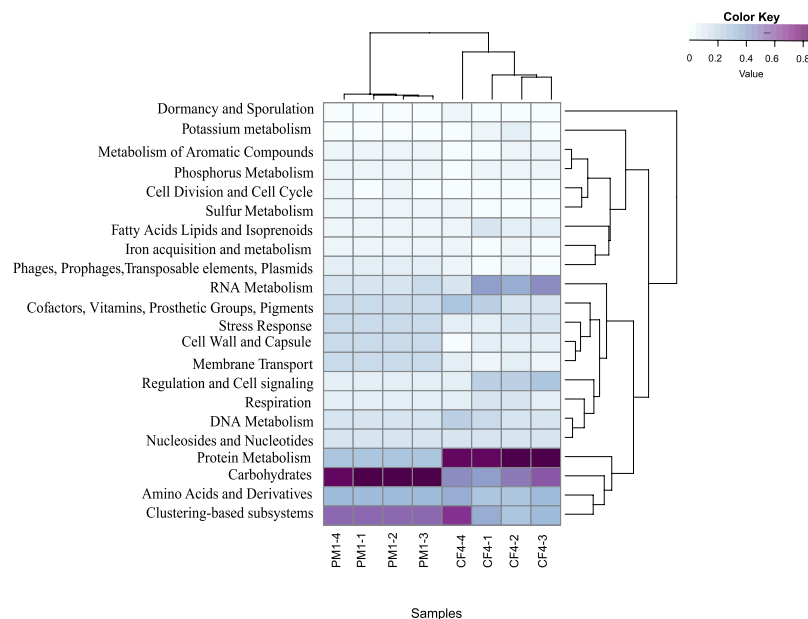
Recently, various strains of *Klebsiella* were tested to determine those with the highest biofilm production and the genes associated with this process<sup>13</sup>. The strain identified with the highest output of biofilm implicated the *RbsA* and *RbsC* genes alongside the quorum sensing molecule Autoinducer 2 (AI-2) and the prophage CP4–57 integrase as putatively involved in the process<sup>13</sup>. These important genes associated with *Klebsiella* were also identified in our study (Table 4). The *RapA* gene identified has been shown to play a role in regulating the *yhcQ* gene that encodes a putative multidrug efflux pump and *yeeZ*, a gene associated with biofilm production<sup>21</sup>. Polyphosphate kinase has been linked with biofilm development, quorum sensing and virulence in *P. aeruginosa* and was also annotated to *Klebsiella* in our study<sup>47,48</sup>.

***Escherichia*.** The surface molecule encoding *wzb* and *wzc* genes, important for capsule assembly and early stage biofilm formation in *K. pneumoniae*<sup>13,16</sup>, were also associated with *Escherichia* in our study (Table 4).

***Enterococcus*.** Similar to genes involved in biofilm formation (the *RbsA*, *SugE*, *ClpX*, *LuxR*, *CspD*) discussed for *Klebsiella* above were associated with *Enterococcus* in our study. *Enterococcus* also exhibited the *galE* gene shown to influence lipopolysaccharide structure, colonisation and biofilm formation<sup>49</sup>.

**Other relevant cellular processes.** Consistent with a more prominent presence of number of species and genes encoding for processes associated with resistance to oxidative stress from bile and biofilm production, patient PM1 also exhibited an overall higher abundance of genes involved in stress response, cell wall and capsule production, cluster-based subsystem activity as well as carbohydrate metabolism compared to patient CF4 (Fig. 4). A potentially enhanced carbohydrate metabolism in PM1 may be attributed to low levels of nutrients in the gallbladder resulting in microbes to metabolise excess biofilm (namely polysaccharides), as reported previously<sup>50</sup>. In contrast, Gram-positive bacteria, mainly associated with gallstones of patient CF4, reveal a higher relative abundance of genes involved in protein metabolism. One hypothesis is that these bacteria are involved in the decomposition of dead human cells associated with the gallstones of this patient (Fig. 4) as inferred from the high relative abundance of human genes compared to PM1. Another explanation for the observed difference in a carbohydrate vs. protein dominated microbial metabolism between the gallstone types is that biofilms of Gram-negative and Gram-positive bacteria differ chemically from each other. EPS produced by Gram-negative bacteria, which predominate in PM1, exhibit anionic properties (attributed to uronic acids) that enable calcium and magnesium ions to bind with polymer strands providing a more tightly bound biofilm architecture<sup>45</sup>. Gram-positive bacteria (mainly in CF4) have been shown to exhibit a more cationic EPS charge, and be composed of teichoic acid mixed with small quantities of protein<sup>45,51</sup>, which may explain a higher relative abundance of genes involved in protein metabolism.





**Figure 4.** Heatmap with the major functional gene categories (acquired from the Subsystems, SEED Database) identified in the gallstones ( $n = 4$ ) of patients PM1 (pigmented) and CF4 (cholesterol). The color key shows the relative abundance of the gene categories in the gallstones. The dendrograms illustrate the relationship between samples showing that the distribution of genes is relatively similar between replicate stones, but greatly differ between the two patients. For example, bacterial genes involved in the carbohydrate vs. protein metabolism were more abundant in gallstones from PM1 vs. CF4.

**Cholesterol analysis.** We conducted compound specific isotopic analysis (CSIA) of the patients carbon ( $\delta^{13}\text{C}$ ) and hydrogen ( $\delta\text{D}$ ) isotopes of cholesterol to ascertain possible dietary or exogenous environmental factors that may be associated with or divergent from the bacteria identified in this study. Individual compounds in a complex mix (i.e. a gallstone made from patient-specific cholesterol/bile mixtures) can have distinct differences in their carbon, hydrogen, oxygen, nitrogen or sulfur isotopic signatures<sup>52</sup>. The marked difference between the isotopic weight of  $\delta^{13}\text{C}$  of an identified compound in two gallstones, for example, can indicate that a different source or mechanism was utilised during the synthesis of the compound<sup>52</sup>. We investigated whether or not the patients' native stones were homogenous and if homogeneity existed between the patients themselves.

The only molecular component identified by GC-MS within the stones of each patient was the Cholest-5-en-3 $\beta$ -ol compound with trace amounts of 5 $\alpha$ -Cholest-7-en-3 $\beta$ -ol. Subsequently, CSIA was conducted on the non-derivatised Cholest-5-en-3 $\beta$ -ol compound of each of the patients' stones (Table 5).

The Cholest-5-en-3 $\beta$ -ol  $\delta^{13}\text{C}$  values obtained for patient PM1 showed a range between  $-24.7$  and  $-25.9\text{‰}$ , and a range between  $-23.1$  and  $-23.7\text{‰}$  for patient CF4, resulting in an approximately  $1\text{‰}$  difference across the four stones analysed for each patient. The Cholest-5-en-3 $\beta$ -ol  $\delta\text{D}$  showed a value range of  $-218$  and  $-231\text{‰}$  for patient PM1, resulting in an approximately  $13\text{‰}$  difference across the four stones analysed for this patient. The Cholest-5-en-3 $\beta$ -ol  $\delta\text{D}$  showed a value range of  $-252$  and  $-254\text{‰}$  for patient CF4, resulting in a negligible  $2\text{‰}$  difference across the four stones analysed for this patient.

Within each patient both the  $\delta^{13}\text{C}$  and  $\delta\text{D}$  values for Cholest-5-en-3 $\beta$ -ol were not significantly different amongst the 4 stones analysed, supporting a common source for Cholest-5-en-3 $\beta$ -ol. Between the two patients the  $\delta^{13}\text{C}$  and  $\delta\text{D}$  values differed only by minor amounts also supporting a common source for Cholest-5-en-3 $\beta$ -ol.

## Conclusions

This pilot study explored taxonomic and functional metagenomics and sterol homogeneity within two patients of diverse backgrounds to elucidate a possible universal factor at play in gallstone pathogenesis and formation. For the first time, functional genes were identified that were associated with bile stress response and biofilm development as possible microbial processes leading to the formation of both pigmented and cholesterol-type gallstones. In the analysed pigmented stones, genes involved in biofilm formation were mainly recovered from clinically pathenogenic *Klebsiella* and *Enterococcus* while bile resistance genes were present also in *Escherichia*, *Shigella*, *Serratia* and *Bacillus*. *Klebsiella* was also present in one of the cholesterol gallstones, while the remaining analysed cholesterol stones showed a predominance of Gram-positive bacteria that were not identified within the pigmented stones. *Klebsiella* was also the only genus to exhibit DMSO respiration, giving it a distinct advantage in the anoxic environment of the human gallbladder. This, in conjunction with being the genus to exhibit the highest number of genes involved in bile stress response, and biofilm formation, may place *Klebsiella* as a major player in gallstone pathogenesis. Further, pigmented stones, predominated by Gram-negative bacteria, revealed a high proportion of genes involved in carbohydrate metabolism, whilst cholesterol stones indicated a profile dominated by protein metabolism. A possible explanation for the observed difference in a carbohydrate

$\delta^{13}\text{C}$ (‰VPDB)				$\delta^{\text{D}}$ (‰VSMOW)			
PM1-1	−24.7(0.1)3	CF4-1	−23.1(0.1)3	PM1-1	−231(0)3	CF4-1	−252(0)3
PM1-2	−25.9(0.1)3	CF4-2	−23.5(0.1)3	PM1-2	−222(1)3	CF4-2	−253(0)3
PM1-3	−24.8(0.4)3	CF4-3	23.3(0)3	PM1-3	−218(3)3	CF4-3	254(0)3
PM1-4	−24.9(0.1)3	CF4-4	23.7(0)3	PM1-4	−221(1)3	CF4-4	254(1)3

**Table 5.** Compound Specific  $\delta^{13}\text{C}$  and  $\delta^{\text{D}}$  results for the patients PM1 and CF4.

vs. protein dominated microbial metabolism between the gallstone types is that biofilms of Gram-negative and Gram-positive bacteria differ chemically from each other resulting with the latter having a higher protein content in the EPS matrix. Fungal and archaeal genes were not detected in both types of stones. Complementary carbon and hydrogen isotopic analyses of cholesterol within the patients' stones revealed homogeneity, suggesting a common diet or cholesterol synthesis pathway that has only a minor influence on microbial composition.

This pilot study provides a framework to study microbial processes that play a potential role in gallstone formation across markedly different types of stones and patient backgrounds. In addition, future studies could also involve metatranscriptomic profiling to ultimately reveal which bacteria are actively expressing genes involved in processes such as bile stress response and biofilm formation that could contribute to the pathogenesis of gallstones.

## Methods

**Sample collection.** Samples were collected whilst the patients were undergoing laproscopic cholecystectomy and were immediately rinsed in sterile saline solution ( $9\text{ g L}^{-1}\text{ NaCl}$ ) and placed in sterile glass containers. The samples were immediately stored at  $-80^{\circ}\text{C}$  until further processing. All patients provided written informed consent upon enrolment to the study. The study was designed with the aid of the National Health and Medical Research Council (NHMRC), according to the guidelines stipulated in the National Statement on Ethical Conduct in Human Research 2007<sup>53</sup>, and the Australian Code for the Responsible Conduct of Research 2007<sup>54</sup>. The study, which includes all associated experiments and methods, was approved by and met the ethical guidelines of the South Metropolitan Health Service Human Research Ethics Committee (HREC Reference: 15–136), the Fiona Stanley Human Research Ethics Committee (Ref: 2015–136), the St John of God Health Care Human Research Ethics Committee (Ref: 1021), and the Curtin University Human Research Ethics Committee (Ref: HR229/2015).

**DNA extraction.** Genomic DNA of 4 gallstones from each patient were obtained from extractions following the procedure described by Haigh and Lee<sup>9</sup>. Inside a HEPA-filtered laminar flow bench, individual gallstones ( $\sim 100\text{ mg}$  each) were pulverized using a heat-sterilized mortar ( $500^{\circ}\text{C}$ , 8 h). 700  $\mu\text{L}$  of 1% SDS solution was added to each pulverized gallstone and incubated under rotation at room temperature for 12 h. Lithium chloride was added (final concentration of 1.5 M) following cell lysis through homogenisation in a FastPrep 96 Instrument (MP Biomedicals LLC, NSW, Australia) (1600 rpm, 60 sec). After centrifugation (5 min, 10,000 rcf) 1 vol of Phenol-Chloroform-isoamylalcohol 25:24:1 (PCI) pH 8 was added to the supernatant, vortexed for 1 min, and centrifuged for 5 min at 10,000 rcf. The PCI extraction was repeated once and 0.4 vol of molecular grade 80 vol% ethanol was added to the aqueous phase. The sample was then transferred to a Spin<sup>TM</sup> Filter and DNA was eluted from the filter using Solution C6 following the guidelines of the PowerSoil DNA Isolation Kit (Mo Bio Laboratories Inc, CA, U.S.A). PCR-inhibiting impurities were completely removed using the OneStep PCR Inhibitor Removal Kit (Zymo Research, U.S.A). The DNA concentration was quantified fluorometrically (NanoDrop 3300 Fluorospectrometer; Thermo Fisher Scientific, MA, U.S.A) using the Quant-iT PicoGreen dsDNA Assay kit (Life Technologies, VIC, Australia).

**Quantitative PCR.** To quantify the amount of bacterial 16S rRNA gene copies, an aliquot of the extracted and purified genomic DNA was subjected to quantitative polymerase chain reaction (qPCR) using general primers<sup>55</sup> targeting the V4 region of bacterial 16S rRNA. All reactions were performed using SYBR Premix Ex Taq (TL1 RNase H Plus) (Takara Bio Inc) in a Realplex quantitative PCR cyclor (Eppendorf) and involved initial denaturing (1 min at  $95^{\circ}\text{C}$ ), followed by 32 cycles including denaturing (5 s at  $95^{\circ}\text{C}$ ), primer annealing (30 sec at  $60^{\circ}\text{C}$ ), primer extension and imaging of newly formed fluorescent (SYBR<sup>®</sup> green I labelled) double-stranded DNA ( $72^{\circ}\text{C}$  for 60 sec). Between  $10^1$  and  $10^6$  copies (10-fold dilution series) of bacterial 16S rRNA were added to reaction mixtures and served as standards during qPCR to calibrate the copy numbers of bacterial 16S rRNA in the gallstone samples.

**Metagenomic library preparation and sequencing.** Metagenomic libraries were prepared using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England BioLabs Inc.) according to manufacturer's instructions. The amplification involved 13–15 cycles. The resulting libraries were concentrated to a volume of 20  $\mu\text{L}$  using Amicon Ultra centrifugal filter units Ultra-0.5 MWCO 30KDa. Gel electrophoresis (2 wt%, 50 min, 120 V) was performed with 10  $\mu\text{L}$  of the concentrated libraries and gel fragments (200–500 bp) were excised and gel purified with the Monarch DNA Gel Extraction Kit (New England BioLabs Inc). The final volume after gel purification for each barcoded library ( $n = 8$ ) was 20  $\mu\text{L}$  and were sent to the Australian Genomic Research Facility (AGRF) in Perth, Western Australia for final quality checking and sequencing. At AGRF, the Illumina HiSeq 2500 platform was used to generate  $2 \times 100\text{-bp}$  pair-end sequencing reads. The HiSeq Control Software (HCS) v2.2.68 and Real Time Analysis (RTA) v1.18.66.3 software performed real-time image analysis and base

calling on the HiSeq instrument computer. The AGRF Illumina bcl2fastq 2.19.0.316 pipeline was used to generate the sequence data.

**Processing of sequence data and bioinformatics.** Approximately 280 million paired-end sequence reads (see Supplementary Table S1) were imported into CLC Genomics Workbench 8.0 (CLC Bio) and trimmed of ambiguous reads to a quality limit of 0.5. Contigs were assembled using the CLC Genomics Workbench paired-end Illumina (*de novo*) read assembler with automatic bubble and word size, length fraction of 0.5, similarity fraction of 0.95, and a minimum contig size cut-off of 300 nucleotides. Contigs were assembled without scaffolding to reduce the formation of chimaeric assemblies. The CLC Genomics Workbench read mapping option was used to map reads onto contigs. ORFs within the contigs were detected using FragGeneScan<sup>8</sup>. Taxonomic assignments of contigs were performed using the NCBI BLASTp software suite against the SEED database of predicted proteins from cultivated microbial genomes with assigned taxonomy. The basis for taxonomic assignment of the ORFs was amino acid similarity of >60% over an alignment length of >50 amino acids to predicted proteins present in the database with an assigned taxonomy<sup>8</sup>.

A matrix showing the relative abundance (average coverage) of annotated ORFs deriving from specific taxa per sample was produced using a python script publicly available online (bitbucket.org/wrf), and was subsequently used for downstream analysis. Heatmaps were performed in R (<http://www.r-project.org/>) using the vegan (<http://vegan.r-forge.r-project.org/>) and the Bioconductor Heatplus (<https://bioconductor.org/biocLite.R>) package. The data was normalized, with the Hellinger function used to produce the taxonomy overview heatmap to show species that may have been obscured by the dominant reads (Fig. 3). The overview of Subsystems, Level 1, cellular processing category annotations were obtained from the SEED database via MG-RAST (Project ID: mgp81110–81111; metagenomics.anl.gov). The Primer-E software package (<http://www.primer-e.com/>) was used to generate principle coordinate analysis (PcoA) plots using the Bray-Curtis distance metric.

**Gas Chromatography Isotope Ratio Mass Spectrometry.** Four gallstones from each patient were individually crushed in heat-sterilized (500 °C, 8 h) mortars. The grounded powder was then extracted *via* sonication (1 h) with dichloromethane (DCM) and methanol (9:1). The extracts were then fractionated by small-scale column liquid chromatography<sup>56</sup>. Approximately 2 mg of the total extract was placed on top of a small column (5 × 0.5 cm i.d.) of activated silica gel (160 °C, 8 h). The first hydrocarbon fraction was eluted with *n*-hexane (2 mL), the second hydrocarbon fraction with DCM in *n*-hexane (1:4, 2 mL), and the more polar fraction with an equal mixture of DCM and methanol (1:1, 2 mL). The fractions were analysed by gas chromatography-mass spectrometry (GC-MS).

The polar fractions (containing Cholest-5-en-3 $\beta$ -ol) were each analysed by compound specific isotope analyses to obtain  $\delta^{13}\text{C}$  and  $\delta\text{D}$  values of Cholest-5-en-3 $\beta$ -ol. The instrument used was a Thermo Delta V Advantage isotope ratio monitoring mass spectrometer (irMS), coupled to a Thermo Trace GC Ultra via a GC Isolink and Conflo IV. The column used was an Agilent DB-5MS Ultra-Inert, 60 m long, 0.25 mm (i.d.), with 0.25  $\mu\text{m}$  film thickness. An aliquot of 1  $\mu\text{L}$  of each fraction was injected into the split/splitless injector in splitless mode, held at 280 °C. The GC oven was increased from 40 to 325 °C at 10 °C/min, then held at 325 °C for 10 min. The carrier gas used was helium held at a constant flow of 1.5 mL/min.

For the carbon isotope analysis, GC column outflow was passed through the GC Isolink combustion reactor (copper oxide / nickel oxide, 1000 °C) to combust hydrocarbons to  $\text{CO}_2$ . For hydrogen isotope analysis, the outflow passed through the high-temperature conversion reactor (graphite-lined, 1420 °C) and was pyrolysed to  $\text{H}_2$ . The  $\text{CO}_2$  /  $\text{H}_2$  passed through the Conflo IV interface to the irMS, which measured  $m/z$  44, 45 and 46 (for  $\text{CO}_2$ ) or  $m/z$  2 and 3 (for  $\text{H}_2$ ). The  $\delta^{13}\text{C}$  and  $\delta\text{D}$  values were calculated from the measured masses by Thermo Isodat software, and calibrated to the VPDB (for  $\text{CO}_2$ ) and VSMOW (for  $\text{H}_2$ ) scales by comparison with a mixture of *n*-alkane standards of known isotopic composition.

**Data availability.** Data are available as raw sequence reads from the NCBI Short Read Archive (SRA) under accession number SRP136827 and as assembled contigs in MG RAST (metagenomics.anl.gov) under accession numbers 4754155.3, 4754325.3, 4754326.3, 4754607.3, 4754608.3, 4754609.3, 4754610.3, 4754611. Fasta files containing the expressed ORFs with signal peptides are available from the authors upon request.

## References

- Maki, T. Pathogenesis of calcium bilirubinate gallstone: Role of *E. coli*, beta glucuronidase and coagulation by inorganic ions, polyelectrolytes, and agitation. *Ann Surg.* **164**, 90–100 (1996).
- Stewart, L., Smith, A. L., Pellegrini, C. A., Motson, R. W. & Way, L. W. Pigment gallstones form as a composite of bacterial micro-colonies and pigment solids. *Ann Surg.* **206**, 242–250 (1987).
- Stewart, L., Ponce, R., Oesterk, A. L., Griffiss, J. M. & Way, L. W. Pigment gallstone pathogenesis: Slime production by biliary bacteria is more important than beta-glucuronidase production. *Gastrointest Surg.* **4**, 547 (2000).
- Stewart, L., Oesterle, A. L., Erdan, I., Griffiss, J. M. & Way, L. W. The pathogenesis of pigment gallstones in western societies: the central role of bacteria. *J Gastrointest Surg.* **6**, 891–904 (2002).
- Cetta, F. The role of bacteria in pigment gallstone disease. *Ann Surg.* **213**, 315–26 (1991).
- Marshall, J. M., Flechtner, A. D., La Perle, K. M. & Gunn, J. S. Visualization of extracellular matrix components within sectioned *Salmonella* biofilms on the surface of human gallstones. *Plos One.* **9**, 1–7 (2014).
- Wu, T. *et al.* Gut microbiota dysbiosis and bacterial community assembly associated with cholesterol gallstones in large-scale study. *BMC Genomics.* **14**, 1–11 (2013).
- Orsi, W. D., Richards, T. A. & Francis, W. R. Predicted microbial secretomes and their target substrates in marine sediment. *Nature Microbiology.* **3**, 32–37 (2018).
- Haigh, W. G. & Lee, S. P. Identification of oxysterols in human bile and pigment gallstones. *Gastroenterology.* **121**, 118–123 (2001).
- Miettinen, T. E. *et al.* Noncholesterol sterols in bile and stones of patients with cholesterol and pigment stones. *Hepatology.* **23**, 274–280 (1996).

11. Shen, H. *et al.* Metagenomic sequencing of bile from gallstone patients to identify different microbial community patterns and novel biliary bacteria. *Sci. Rep.* **5**, 17450 (2015).
12. Swidinski, A. & Lee, S. P. The role of bacteria in gallstone pathogenesis. *Frontiers in Bioscience*. **6**, e93–103 (2001).
13. Landini, P. Cross-talk mechanisms in biofilm formation and responses to environmental and physiological stress in *Escherichia coli*. *Res Microbiol.* **160**, 259–66 (2009).
14. Iwase, T., Ogura, Y., Hayashi, T. & Mizunoe, Y. Complete genome sequence of *Klebsiella pneumoniae* YH43. *Genome Announc.* **4**, 00242–16 (2016).
15. Klappenbach, J. A., Dunbar, J. M. & Schmidt, T. M. rRNA operon copy number reflects ecological strategies of Bacteria. *Applied and Environmental Microbiology*. **66**, 1328–1333 (2000).
16. Begley, M., Gahan, C. G. & Hill, C. The interaction between bacteria and bile. *FEMS Microbiology Reviews*. **29**, 625–651 (2005).
17. Wu, M., Lin, T., Hsieh, P., Yang, H. & Wang, J. Isolation of genes involved in biofilm formation of a *Klebsiella pneumoniae* strain causing Pyogenic Liver Abscess. *Plos One*. **6**, 1–11 (2011).
18. Jeticia, R. *et al.* Survival of the Fittest: How bacterial pathogens utilize bile to enhance infection. *Clinical Microbiology Reviews*. **29**, 819–836 (2016).
19. Tortora, G. J. & Derrickson, B. *Principles of anatomy & physiology*. 944–950 (John Wiley & Sons, 2014).
20. Walsh, C. Molecular mechanisms that confer antibacterial drug resistance. *Nature*. **406**, 775–781 (2000).
21. Lebeaux, D., Ghigo, J. M. & Beloin, C. Biofilm-Related Infections: Bridging the gap between clinical management and fundamental aspects of recalcitrance toward antibiotics. *Microbiol Mol Biol Rev.* **78**, 510–543 (2014).
22. Sulavik, M. C., Dazer, M. & Miller, P. F. The *Salmonella typhimurium* mar locus: molecular and genetic analyses and assessment of its role in virulence. *J. Bacteriol.* **179**, 1857–1866 (1997).
23. Prouty, A. M., Brodsky, I. E., Falkow, S. & Gunn, J. S. Bile-salt-mediated induction of antimicrobial and bile resistance in *Salmonella typhimurium*. *Microbiology*. **150**, 775–783 (2004).
24. Bina, J. E. & Mekalanos, J. J. *Vibrio cholerae* tolC is required for bile resistance and colonization. *Infect. Immun.* **69**, 4681–4685 (2001).
25. Thanassi, D. G., Cheng, L. W. & Nikaido, H. Active efflux of bile salts by *Escherichia coli*. *J. Bacteriol.* **179**, 2512–2518 (1997).
26. Baranova, N. & Nikaido, H. The BaeSR two-component system activates transcription of the yegMNOB (mdtABCD) transporter gene cluster in *Escherichia coli* and increases its resistance to novobiocin and deoxycholate. *J. Bacteriol.* **184**, 4168–4176 (2002).
27. Lin, J., Sahin, O., Overbye Michel, L. & Zhang, Q. Critical role of multidrug efflux pump CmeABC in bile resistance and *in vivo* colonization of *Campylobacter jejuni*. *Infect. Immun.* **71**, 4250–4259 (2003).
28. Heithoff, D. M. *et al.* *Salmonella* DNA adenine methylase mutants confer cross protective immunity. *Infect. Immun.* **69**, 6725–6730 (2001).
29. Le Breton, Y., Maze, A., Hartke, A., Lemarini, S. & Auffray, Y. & Rince', A. Isolation and characterization of bile salts sensitive mutants of *Enterococcus faecalis*. *Curr. Microbiol.* **45**, 434–439 (2002).
30. Rincé, A., Giard, J. C., Pichereau, V., Flahaut, S. & Auffray, Y. Identification and characterization of gsp65, an organic hydroperoxide resistance (ohr) gene encoding a general stressprotein in *Enterococcus faecalis*. *J. Bacteriol.* **183**, 1482–1489 (2001).
31. Leverrier, P., Vissers, J. P. C., Rouault, A., Boyaval, P. & Jan, G. Mass spectrometry proteomic analysis of stress adaptation reveals both common and distinct response pathways in *Propionibacterium freudenreichii*. *Arch. Microbiol.* **181**, 215–230 (2004).
32. Hu, Y., Lu, P., Zhang, Y., Li, L. & Chen, S. Characterization of an aspartate-dependent acid survival system in *Yersinia pseudotuberculosis*. *FEBS Letters* **584**, 2311–2314 (2010).
33. Smart, R. C. & Hodgson, E. *Molecular and Biochemical Toxicology*, John Wiley & Sons, New Jersey (Section 18.6) (2013).
34. Dal Santo *et al.* Stress-induced GSK3 regulates the redox stress response by phosphorylating glucose-6-phosphate dehydrogenase in *Arabidopsis*. *Plant Cell*. **24**, 3380–92 (2012).
35. Arese, P., Gallo, V., Pantaleo, A. & Turrini, F. Life and death of Glucose-6-Phosphate Dehydrogenase (G6PD) deficient erythrocytes – Role of redox stress and Band 3 modifications. *Transfusion Medicine and Hemotherapy*. **39**, 328–334 (2012).
36. Lu, A. L., Cuipa, M. J., Ip, M. S. & Shanabruch, W. G. Specific A/G-to-C/G mismatch repair in *Salmonella typhimurium* LT2 requires the mutB gene product. *J. Bacteriol.* **172**, 1232–1240 (1990).
37. Oliveira L. C. *et al.* Analyses of the probiotic property and stress resistance-related genes of *Lactococcus lactis* subsp. *lactis* NCDO 2118 through comparative genomics and *in vitro* assays. *Plos One* **12** (2017).
38. Gralnick, J. A., Vali, H., Lies, D. P. & Newman, D. K. Extracellular respiration of dimethyl sulfoxide by *Shewanella oneidensis* strain MR-1. *PNAS*. **103**, 4669–4674 (2006).
39. Chowdhury, R., Gautam, K. S. & Jyotirmoy, D. Stress response in pathogenic bacteria. *J. Biosci.* **21**, 149–160 (1996).
40. Lee, H. S. *et al.* Mechanism of regulation of 8-hydroxyguanine endonuclease by oxidative stress: roles of *fmr*, *arcA*, and *fur*. *Free Radic. Biol. Med.* **24**, 1193–1201 (1998).
41. Fink, R. C. *et al.* FNR is a global regulator of virulence and anaerobic metabolism in *Salmonella enterica* Serovar *Typhimurium* (ATCC 14028s). *J. Bacteriol.* **189**, 2262–2273 (2007).
42. Poole, K. *Advances in Microbial Physiology*, Volume 58 (ed. Poole, K.) 84–94 (Elsevier, 2011).
43. van Velkinburgh, J. C. & Gunn, J. S. PhoP-PhoQ-regulated loci are required for enhanced bile resistance in *Salmonella* spp. *Infect Immun.* **67**, 1614–22 (1999).
44. Ito, A., Taniuchi, A., May, T., Kawata, K. & Okabe, S. Increased antibiotic resistance of *Escherichia coli* in mature biofilms. *Applied and environmental microbiology* **75**, 4093–4100 (2009).
45. Donlan, R. M. Biofilms: Microbial Life on Surfaces. *Emerging Infectious Diseases*. **8**, 881–890 (2002).
46. Hentzer, M. & Givskov, M. J. Clin. Pharmacological inhibition of quorum sensing for the treatment of chronic bacterial infections. *J. Clin. Invest.* **112**, 1300–1307 (2003).
47. Bandeira, M., Borges, V., Gomes, J. P., Duarte, A. & Jordao, L. Insights on *Klebsiella pneumoniae* biofilms assembled on different surfaces using phenotypic and genotypic approaches. *Microorganisms*. **5**, 2–16 (2017).
48. Rashid *et al.* Polyphosphate kinase is essential for biofilm development, quorum sensing, and virulence of *Pseudomonas aeruginosa*. *Proc Natl Acad Sci USA* **97**, 9636–41 (2000).
49. Nesper, J. *et al.* Characterization of *Vibrio cholerae* 01 El tor galU and galE mutants: influence on lipopolysaccharide structure colonization, and biofilm formation. *Infect. Immun.* **69**, 435–445 (2001).
50. Bhatt, V. D., Dande, S. S., Patil, N. V. & Chaitanya, G. J. Molecular analysis of the bacterial microbiome in the forestomach fluid from the dromedary camel (*Camelus dromedarius*). *Mol Biol Rep.* **40**, 3363–3371 (2013).
51. Hussain, M., Wilcox, M. H. & White, P. J. The slime of coagulase-negative-staphylococci: biochemistry and relation to adherence. *FEMS Microbiol Rev.* **104**, 191–208 (1993).
52. Mancini, S. A., Lacrampe-Couloume, G. & Lollar, B. S. Source differentiation for benzene and chlorobenzene groundwater contamination: A field application of stable carbon and hydrogen isotope analyses. *Environmental Forensics*. **9**, 177–186 (2008).
53. National Health and Medical Research Council (NHMRC), 2007a, Australian Code for the Responsible Conduct of Research, <https://www.nhmrc.gov.au/guidelines-publications/r39> (2007).
54. National Health and Medical Research Council (NHMRC), 2007b, National Statement on Ethical Conduct in Human Research, <https://www.nhmrc.gov.au/book/national-statement-ethical-conduct-human-research> (2007).
55. Caporaso *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME Journal*. **6**, 1621–1624 (2012).

56. Bastow, T. P., van Aarssen, B. G. K. & Lang, D. Rapid small-scale separation of saturate, aromatic and polar components in petroleum. *Organic Geochemistry*. **38**, 1235–1250 (2007).
57. Perez, M. J. & Briz, O. Bile-acid-induced cell injury and protection. *World Journal of Gastroenterology*. **15**, 1677–1689 (2009).

## Acknowledgements

The Fiona Stanley and St John of God, Murdoch hospitals are thanked for facilitating patient consent, sample collection and ethics approval for the study. The authors would like to acknowledge WA-OIGC, TIGeR, Curtin University, and the contribution of an Australian Government Research Training Program Scholarship in supporting this research. K.G. thanks the Australian Research Council for infrastructure support (LE110100119).

## Author Contributions

S.H.K., M.J.L.C. and K.G. designed the experiments. S.H.K., M.J.L.C., W.D.O. and K.G. carried out the analyses and data interpretation. S.H.K. and M.J.L.C. wrote the main manuscript text with contributions from K.G., W.D.O. and M.B.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-29571-8>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018